PTO/SB/21 (12-97)
Approved for use through 9/30/00. OMB 0651-0031
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paper Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Application Number	09/553,956
Filing Date	April 21, 2000
In re Application of:	Thomas RUNKLER et al.
Group Art Unit	2172
Examiner Name	Pham, H.
Attorney Docket Number	50277-0452
Client Docket Number	OID-1999-038-01
Total Number of Pages in This Submission	29

ENCLOSURES ***(check all that apply)***

- [x] Appeal Brief Fee $500
  - [x] Credit Card Transmittal Attached
- [ ] Amendment / Response
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Response to Missing Parts/ Incomplete Application
  - [ ] Response to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Assignment Papers *(for an Application)*
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition Routing Slip (PTO/SB/69) and Accompanying Petition
- [ ] To Convert a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Small Entity Statement
- [ ] Request of Refund
- [ ] After Allowance Communication to Group
- [ ] Appeal Communication to Board of Appeals and Interferences
- [x] Appeal Communication to Group *(Appeal Notice, Brief, Reply Brief)*
- [ ] Proprietary Information
- [ ] Status Letter
- [ ] Additional Enclosure(s) *(please identify below)*:

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm *or* Individual name	DITTHAVONG & CARLSON, P.C. Stephen C. Carlson, Reg. No. 39929
Signature	
Date	January 19, 2005

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Assistant Commissioner for Patents, Alexandria, VA 22313-1450 on this date:

Type or printed name	Linda V. Wiley		
Signature		Date	January 19, 2005

09/553,956 *Patent*

CANCELLED

JAN 18 2005

JAN 21 2005

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of:	Thomas RUNKLER et al.	Conf. No.:	7423
Application No.:	09/553,956	Group Art Unit:	2172
Filed:	April 21, 2000	Examiner:	Pham, H.
Attorney Docket:	50277-0452		
Client Docket:	OID-1999-038-01		

For: SYSTEM AND METHOD FOR GENERATING DECISION TREES

APPEAL BRIEF

Honorable Commissioner for Patents
Alexandria, VA 22313-1450

Dear Sir:

This Appeal Brief is submitted in support of the Notice of Appeal dated November 18, 2004.

I. REAL PARTY IN INTEREST

Oracle International Corporation is the real party in interest.

II. RELATED APPEALS AND INTERFERENCES

Appellants are unaware of any related appeals and interferences.

01/24/2005 ZJUHAR1 00000067 09553956
01 FC:1402 500.00 OP

01/24/2005 ZJUHAR1 00000020 09553956
~~01 FC:1402~~ ~~500.00 OP~~

Void date: ~~01/24/2005 ZJUHAR1~~ 00000020 ~~09553956~~
~~01/24/2005 ZJUHAR1~~
~~01 FC:1402~~ ~~500.00 OP~~

III. STATUS OF THE CLAIMS

Claims 1-8, 10, 12-25, 27, and 29-36 are pending in this appeal, in which claims 9, 11, 26, and 28 have earlier been canceled. Claims 7-8 and 24-25 are allowed and claims 15 and 32 are indicated as allowable. This appeal is therefore taken from the final rejection of claims 1-6, 10, 12-14, 16-23, 27, 29-31, and 33-36 on May 20, 2004.

IV. STATUS OF AMENDMENTS

No amendment to the claims has been filed subsequent to the final rejection dated May 20, 2004.

V. SUMMARY OF THE INVENTION

The present invention addresses problems associated with generating decision trees (Specification, p. 1: 4-5). A common conventional approach to build decision trees is known as "Induction of Decision Trees" or ID3, which is a recursive algorithm that starts with a set of training objects that belong to a set of predefined classes. If all the objects belong to a single class, then there is no decision to make and a leaf node is created and labeled with the class. Otherwise, a branch node is created and the attribute with the highest "information gain" is selected if that attribute were used to discriminate objects at the branch node. The information gain is calculated by finding the average entropy of each attribute. (Specification, p. 3:11-17)

A problem with conventional decision trees such as those produced by ID3 is that such decision trees are rigid, inflexible, and brittle (Specification, p. 3:18-19). FID3 attempts to combine fuzzy logic with classical, crisp decision trees. In FID3, the user defines the membership functions in each of the predefined classes for all of the training data. Each membership function can serve as an arc label of a fuzzy decision. As in ID3, FID3 generates

its decision tree by maximizing information gains. The decision of the fuzzy decision tree is also a fuzzy variable, indicating the memberships of a tested object in each of the possible classifications (Specification, p. 4:10-15). One disadvantage with FID3 is that the membership functions in each of the attributes for all of the training data must be specified beforehand by the user. For data with a high number of attributes or dimensions, however, determining the membership functions is typically a difficult task, requiring intensive involvement by experts. In addition, the fuzzy sets themselves may not even be known beforehand and require further investigation. (Specification, pp. 4:23–5:2)

The present invention solves these problems by providing a data analysis technique that is capable of handling real-world or "fuzzy" data in a flexible manner, and a technique in which the groupings of the data or other *a priori* information, such as fuzzy membership functions, need not be supplied beforehand, and data can be dynamically clustered while a decision tree is generated. In one embodiment, the data are clustered using a fuzzy clustering analysis, which generates the membership functions on the fly, without requiring the user to predefine sets or calculate the membership functions beforehand. (Specification, p. 6:2-6)

As an example, for a two-dimensional data set, the data in each dimension (*e.g.* x and y) are clustered and a partition coefficient, which quantifies the goodness of the clustering, is computed for each dimension as a measure of cluster validity or how well separated the clusters are. In one implementation, fuzzy c-means clustering may be employed, but other forms of fuzzy clustering such as fuzzy k-means may be employed. In a fuzzy c-means clustering approach, the objective function J_{FCM} is minimized for a given number of clusters c. Thus, fuzzy clustering is performed for several different clustering numbers (for example, up to a c = 4 clusters) leading to partitions $U_c^{(l)}$ and a partition coefficient $PC(U_c^{(l)})$ is calculated.

In one embodiment, PC($U_c^{(l)}$) is calculated for $c > 1$ as follows (Specification, p. 15:3-13):

$$\mathrm{PC}(U_c^{(l)}) = \tfrac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ikc}^{(l)})^2 ,$$

where the FCM model may be defined as the minimization of the objective function J_{FCM} for a given data set $X = \{ x_i \}$, $i \in 1..n$ with dimensions $l \in 1..p$, and a fuzziness parameter $m \in (1, \infty)$:

$$J_{FCM}(U,V;X) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^{m}\left|x_k - v_i\right|^2, \quad (1)$$

where $U = \{ u_{ik} \}$, $V = \{ v_i \}$, $u_{ik} \in [0, 1]$ is the membership of x_k in the i^{th} cluster of c clusters, $i \in 1..c$, $k \in 1..n$, with $\sum u_{ik} = 1$, for all $k \in 1..n$, and v_i is the center of the i^{th} cluster, $i \in 1..c$, and m is typically 2. (Specification, p. 13:1-7)

A decision tree clustering procedure consistent with the present invention employs a unified approach to extracting both the decision tree and the (crisp or fuzzy) clusters. The decision tree is built by subsequent clustering of single dimensions or features, and the choice of the winning separation is based on cluster validity. In one embodiment, the clustering employs a fuzzy c-means (FCM) model and the partition coefficient (PC) to determine the selected separations. Use of the partition coefficient as the cluster validity measure produces results that are good or optimal with respect to cluster separability. Other optimality conditions, however, can be incorporated by choosing other validity measures, and clustering models other than FCM can be employed for generating decisions trees. For example, the use of a hard c-means (HCM) model instead of FCM, for example, leads to crisp decision trees. (Specification, pp. 17:19–18:3)

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Whether claims 1-3 and 18-20 are obvious under 35 U.S.C. § 103(a) based on *Rastogi et al.* (U.S. 6,247,016) in view of *Shimoji et al.* ("Data Clustering with Entropical Scheduling").

Whether claims 1-5 and 18-22 are obvious under 35 U.S.C. § 103(a) based on *Janikow* ("Fuzzy Decision Trees: Issues and Methods") and *Choe et al.* ("On the Optimal Choice of Parameters in a Fuzzy C-Means Algorithm").

Whether claims 4 and 21 are obvious under 35 U.S.C. § 103(a) based on *Rastogi et al.*, *Shimoji et al.*, and *Background.*

Whether claims 5 and 22 are obvious under 35 U.S.C. § 103(a) based on *Rastogi et al.*, *Shimoji et al.*, *Background*, and *Hall et al.* ("Generating Fuzzy Rules from Data").

Whether claims 6 and 23 are obvious under 35 U.S.C. § 103(a) based on *Rastogi et al.*, *Shimoji et al.*, and *Shafer et al.* ("SPRINT: A Scalable Parallel Classifier for Data Mining" 1996).

Whether claims 6 and 23 are obvious under 35 U.S.C. § 103(a) based on *Janikow*, *Choe et al.*, and *Shafer et al.*

Whether claims 10, 12, 16, 27, 29, and 33 are obvious under 35 U.S.C. § 103(a) based on *Janikow*.

Whether claims 13 and 30 are obvious under 35 U.S.C. § 103(a) based on *Janikow* and *Choe et al.*

Whether claims 14 and 31 are obvious under 35 U.S.C. § 103(a) based on *Janikow* and *Shafer et al.*

Whether claims 17 and 34-36 are anticipated under 35 U.S.C § 102(e) by the *Background* section.

VII. ARGUMENT

A. CLAIMS 17 AND 34-36 ARE NOT ANTICIPATED OVER THE *BACKGROUND OF THE INVENTION.*

To anticipate a patent claim, every element and limitation of the claimed invention must be found in a single prior art reference, arranged as in the claim. *Karsten Mfg. Corp. v. Cleveland Golf Co.*, 242 F.3d 1376, 1383, 58 USPQ2d 1286, 1291 (Fed. Cir. 2001); *Scripps Clinic & Research Foundation v. Genentech, Inc.*, 927 F.2d 1565, 1576, 18 USPQ2d 1001, 1010 (Fed. Cir. 1991).

Reversal of the rejection of claims 17 and 34-36 is respectfully requested because claims 17 are 34-36 are not anticipated by anything admitted in the *Background.* As a preliminary matter, the Manual of Patent Examining Procedure § 608.01(c) states that the purpose of the Background section is as follows: "Where applicable, the problems involved in the prior art or **other information** disclosed which are solved by the applicant's invention should be indicated" (emphasis added). Accordingly, unless something is explicitly stated to be prior art, the mere inclusion of subject matter in the Background section is not sufficient by itself to be an admission to be prior art. The *Background* does not admit that the subject matter of claims 17 and 34-36 are in the prior art. In fact, the *Background* is silent about various features of the claims. For example, independent claim 17 recites:

> 17. (Original) A method for generating a decision tree for a plurality of data characterized by a plurality of features, comprising:
>
> performing a plurality of fuzzy cluster analyses along each of the features to calculate a maximal partition coefficient and a corresponding set of one or more fuzzy clusters, said maximal partition coefficient corresponding to one of the features;
>
> selecting the one of the features corresponding to the **maximal partition coefficient**; and
>
> building the decision tree based on the corresponding set of one or more fuzzy clusters.

Independent claim 35 also recites "selecting the one of the features corresponding to the **maximal partition coefficient**," but a maximal partition coefficient is not to be found in the *Background*.

The Examiner's reasoning is predicated on the mistaken assumption that a maximal partition coefficient can be equated to a maximum information gain (Office Action of May 20, 2004, p. 16, emphasis original):

> As seen, μ_{young} (X_i) and μ_{old} (X_i) as a plurality of fuzzy cluster analyses is performed along each of the age features to calculate the highest information gain corresponding to one of the features as ***maximum partition coefficient*** and for two fuzzy sets Young and Old, then the attribute with the highest information gain is selected to discriminate objects at the branch node to build the decision tree based on two fuzzy sets Young and Old.

However, the *Background* merely states at p. 4:13-14 that "As in ID3, FID3 generates its decision trees by maximizing information gains." There is no support in the *Background*, admitted or otherwise, for the Examiner's glossing of "highest information gain" as a "maximum partition coefficient."

The basis for the Examiner's highly unusual understanding appears to be a phrase in the *Detailed Description* of the Specification, p. 15:4, that explains a property of the partition coefficient as "which quantifies the goodness of the clustering" (Office Action, p. 3). This statement in the *Detailed Description*, however, is clearly not found in the *Background* or admitted prior art. Furthermore, quantifying the goodness of the clusters does not mean that any number that might have some connection to fuzzy clustering must be a partition coefficient.

Even with the Examiner's excessively broad construction of "maximum partition coefficient" to be any kind of fuzzy-clustering number, the rejection's reasoning still fails. Both the non-fuzzy ID3 and the fuzzy FID3 generate their trees by maximizing information gains (*Background*, p.4:13-14, quoted above). According to the *Background*, maximizing information

gains is independent of fuzzy clustering, especially since it describes a non-fuzzy methodology, ID3. This can be seen in a detailed discussion of information gain found in *Janikow*, at pp. 5:2–6:1, in its outline of the ID3 partitioning algorithm:

> The root of the decision tree contains all training examples. It represents the whole description space since no restrictions are imposed. Each node is recursively split by partitioning its examples. A node becomes a leaf when either its samples come from a unique class or when all attributes are used on the path. When it is decided to further split the node, one of the remaining attributes (i.e., not appearing on the current path) is selected. Domain values of that attribute are used to generate conditions leading to child nodes. The examples present in the node being split are partitioned into child nodes according to their matches to those conditions. One of the most popular attribute selection mechanisms is one that maximizes information gain [25]. This mechanism, outlined below, is computationally simple as it assumes independence of attributes.
>
> 1) Compute the information content at node *N*, given by $I^N = -\sum_{i=1}^{|C|}(p_i \cdot \log p_i)$, where *C* is the set of decisions, and p_i is the probability that a training example in the node represents class *i*.
>
> 2) For each attribute a_i not appearing on the path to *N* and for each of its domain values a_{ij}, compute the information content $I^{N|a_{ij}}$ in a child node restricted by the additional condition $a_i = a_{ij}$.
>
> 3) Select the attribute a_i maximizing the information gain $I^N - \sum_j^{|D_i|}(w_j \cdot I^{N|a_{ij}})$, where w_j is the relative weight of examples in *N*, and D_i is the symbolic domain of the attribute.
>
> 4) Split the node using the selected attribute.

For these reasons, one of ordinary skill in the art would not understand, based either on the *Background* or the prior art, that either ID3 or FID3 builds their decision trees using a maximal partition coefficient. In fact, such a person of ordinary skill would not even equate a maximal partition coefficient with a maximum information gain. Well-settled case law holds that the words of a claim must be read as they would be interpreted by those of ordinary skill in the art. *In re Baker Hughes Inc.,* 215 F.3d 1297, 55 USPQ2d 1149 (Fed. Cir. 2000); *In re Morris*, 127 F.3d 1048, 1054, 44 USPQ2d 1023, 1027 (Fed. Cir. 1997). "Although the PTO

must give claims their broadest reasonable interpretation, this interpretation must be consistent with the one that those skilled in the art would reach." *In re Cortright,* 165 F.3d 1353, 1369, 49 USPQ2d 1464, 1465 (Fed. Cir. 1999).

Accordingly, the anticipation rejection of claims 17 and 34-36 over the *Background* is inconsistent with how a person of ordinary skill in the art would understand either maximal partition coefficient or maximum information gain and must be reversed.

B. CLAIMS 1-6, 10, 12-14, 16, 18-23, 27, 29-31, AND 33 ARE NOT OBVIOUS OVER *JANIKOW* AND OTHER APPLIED ART.

The initial burden of establishing a *prima facie* basis to deny patentability to a claimed invention under any statutory provision always rests upon the Examiner. *In re Mayne*, 104 F.3d 1339, 41 USPQ2d 1451 (Fed .Cir. 1997); *In re Deuel*, 51 F.3d 1552, 34 USPQ2d 1210 (Fed. Cir. 1995); *In re Bell*, 991 F.2d 781, 26 USPQ2d 1529 (Fed. Cir. 1993); *In re Oetiker*, 977 F.2d 1443, 24 USPQ2d 1443 (Fed. Cir. 1992). In rejecting a claim under 35 U.S.C. § 103, the Examiner is required to provide a factual basis to support the obviousness conclusion. *In re Warner*, 379 F.2d 1011, 154 USPQ 173 (CCPA 1967); *In re Lunsford*, 357 F.2d 385, 148 USPQ 721 (CCPA 1966); *In re Freed*, 425 F.2d 785, 165 USPQ 570 (CCPA 1970).

1. *Janikow* does not suggest calculating "partition coefficients based on membership functions of the data" as recited in claims 10, 12, 16, 27, 29, and 33.

With regard to claims 10, 12, 16, 27, 29, and 33, the rejection over *Janikow* should also be reversed because *Janikow* does not suggest the features of the claims. For example, independent claim 10 recites:

10. (Previously Presented) A method for generating a decision tree for a plurality of data characterized by a plurality of features, comprising:

performing a plurality of cluster analyses along each of the features to calculate a plurality of respective partition coefficients based on membership functions of the data for one or more clusters in respective said cluster analyses;

selecting the one of the features **corresponding to a maximal partition coefficient from among the partition coefficients**;

subdividing the data into one or more groups **based on the selected feature**; and

building the decision tree based on the one or more groups.

Independent claim 27 also recites the selecting and subdividing features.

The Office Action of May 21, 2004, p, 13, admits that "Janikow does not explicitly teach the G^R_{Inc} and G^R_{Emp} as the partition coefficients." Without any teaching of a partition coefficients, there is nothing in *Janikow* to teach or other suggest the following step of selecting and subdividing which recites the "partition coefficients," nor for that matter the next step of subdividing that is "based on the selected feature." In fact, the Examiner recognizes that *Janikow* uses another measure to split the node, viz., "to calculate a plurality of information gain to the split the node." As explained above in Section VII. A., one of ordinary skill in the art would not confuse information gain with a partition coefficient.

Recognizing *Janikow*'s deficiency, the Examiner contends that it would have been obvious "to modify the Janikow method by using function f_2 as the membership function ... in order to split a node" (p. 13). However, *Janikow*, p. 9, expressly teaches against just such a modification: "To define the decision procedure, we must define f_0, f_1, f_2, f_3 for dealing with samples presented for classification. These operators may **differ from those used for tree-building**—let us denote them g_0, g_1, g_2, g_3." Thus, *Janikow* discloses a distinction between classification functions, e.g. f_2, and tree building functions, e.g. g_2, and one of ordinary skill in

the art would **not** be motivated to disregard *Janikow*'s distinctions and principle of operation when making modifications of its method.

Obviousness rejections require some evidence in the prior art of a teaching, motivation, or suggestion to combine and modify the prior art references. See, e.g., *McGinley v. Franklin Sports, Inc.*, 262 F.3d 1339, 1351-52, 60 USPQ2d 1001, 1008 (Fed. Cir. 2001); *Brown & Williamson Tobacco Corp. v. Philip Morris Inc.*, 229 F.3d 1120, 1124-25, 56 USPQ2d 1456, 1459 (Fed. Cir. 2000); *In re Dembiczak*, 175 F.3d 994, 999, 50 USPQ2d 1614, 1617 (Fed. Cir. 1999).It is improper to combine references where the references teach away from their combination. *In re Grasselli*, 713 F.2d 731, 218 USPQ 769 (Fed. Cir. 1983). A prior art reference must be considered in its entirety including portions that would lead away from the claimed invention. *W.L. Gore & Associates, Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 USPQ 303 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 851 (1984).

2. Claims 1-5 and 18-22 are not obvious over *Janikow* and *Choe et al.* because *Janikow* teaches against the invention recited therein.

The rejection of claims 1-5 and 18-22 is also infirm over *Janikow* in view of *Choe et al.* because *Janikow* teaches against the invention recited in claims 1-5 and 18-22. In particular, independent claim 1 recites:

> 1. (Previously Presented) A method for refining a node of a decision tree associated with a plurality of data characterized by a plurality of features, comprising:
>
> selecting a feature from among the features characterizing the data associated with the node;
>
> **performing a cluster analysis along the selected feature** to group the data into one or more clusters based on distances between the data and respective one or more centers of the one or more clusters;
>
> constructing one or more arcs of the decision tree at the node respectively for each of the one or more clusters;

projecting the data in each of the clusters, wherein the projected data are characterized by the plurality of the features but for the selected feature; and

recursively performing the steps of selecting a feature and **performing the cluster analysis** on the projected data in each of the clusters.

However, *Janikow* does not show "recursively ... performing the cluster analysis." The Examiner's rejection, which merely cites pp. 7-9 without explanation, is inadequate, since *Janikow* discloses a distinction between classification functions, e.g. f_2, and tree building functions, e.g. g_2. In fact, by keeping classification and tree building distinct, *Janikow* teaches against "recursively ... performing the cluster analysis" in general and the proposed modification of *Janikow* to use *Choe et al.*'s classification system. Because of this distinction, *Janikow* actually teaches against using any classification function in *Choe et al.* for tree building (cf. claims 1 and 18: "constructing one or more arcs of the decision tree").

5. CLAIMS 6 AND 23 ARE NOT RENDERED OBVIOUS BY *JANIKOW, CHOE ET AL.* AND *SHAFER ET AL.* BECAUSE THE REFERENCES FAIL TO ADDITIONALLY TEACH "PERFORMING A HARD CLUSTER ANALYSIS ."

The Examiner correctly acknowledges that *Janikow* and *Choe et al.* fail to disclose "the step of performing the cluster analysis includes the step of performing a hard cluster analysis," but contends that *Shafer et al.* "teaches a method of forming a decision tree by performing a hard cluster analysis," citing pp. 544-550, "especially Abstract and Introduction pages 544-545." (Office Action, p. 12) However, *Shafer et al.*, directed to a scalable parallel classifier for data mining (per title), discusses only "classes" of data and partitions of the data (e.g., p. 546, left column), and makes no mention of any "cluster analysis," much less "performing a hard cluster analysis." Furthermore, *Shafer et al.*'s different classification function does not undo *Janikow*'s teaching against the invention in claim 1, upon which claim 6 depends.

Therefore, the obviousness rejection of claims 6 and 23 based on *Janikow*, *Choe et al.*, and *Shafer et al.* should also be reversed.

3. CLAIMS 13 AND 30 ARE NOT RENDERED OBVIOUS BY *JANIKOW* AND *CHOE ET AL.*

The rejection of claims 13 and 30 based on *Janikow* and *Choe et al.* should also be reversed. Claim 13 dependent on claim 10. Since *Janikow*'s separation of classification and tree-building teaches against claim 10, *Choe et al.*'s different classification function does not undo *Janikow*'s teaching against.

4. CLAIMS 14 AND 31 ARE NOT RENDERED OBVIOUS BY *JANIKOW* AND *SHAFER ET AL.*

The rejection of claims 14 and 31 based on *Janikow* and *Shafer et al.* should also be reversed. *Janikow*'s separation of classification and tree-building teaches against claim 10, upon which claim 14 depends, and *Shafer et al.*'s different classification function does not undo *Janikow*'s teaching against.

C. CLAIMS 1-6 AND 18-23 ARE NOT OBVIOUS OVER *RASTOGI ET AL.* AND *SHIMOJI ET AL.*

1. CLAIMS 1-3 AND 18-22 ARE NOT RENDERED OBVIOUS BY *RASTOGI ET AL.* AND *SHIMOJI ET AL.*

The rejection of claims 1-3 and 18-22 based on *Rastogi et al.* in view of *Shimoji et al.* should be reversed because *Rastogi et al.* in view of *Shimoji et al.* fail to disclose the limitations of these claims. For example, independent claims 1 and 18 recite: "performing a cluster analysis <u>along the selected feature</u> to group the data into one or more clusters based on distances between the data and respective one or more centers of the one or more clusters."

This limitation is not shown in *Rastogi et al.* Rather, *Rastogi et al.* is directed to a decision tree classifier with integrated building and pruning phases (Title). *Rastogi et al.* involves sample records having multiple attributes, the sample records being identified or "tagged" with a special classifying attribute which indicates a **class** to which the record belongs. For example, as shown in FIG. 1, a training set has sample records identifying the salary level (continuous attributes) and education level (categorical attributes) of a group of applicants for loan approval. Each record is tagged with either an "accept" classifying attribute or a "reject" classifying attribute, depending upon the parameters for acceptance or rejection set by the user of the database (col. 2:33-49). *Rastogi et al.* discloses that its "tree is built breadth-first by recursively partitioning the data until each partition is pure" (col. 3:40-41). *Rastogi et al.* then describes two conditions for splitting the data: if the data A is numeric, then the split is of the form $A < v$, and if data A is categorical, then the split is of the form $A \in V$. Then, *Rastogi et al.* chooses the "split with the least entropy" (col. 4:38) and is therefore maximizing information gains.

Nowhere does *Rastogi et al.* describe "cluster analysis" or even a split based on any type of cluster analysis. In fact, *Rastogi et al.* nowhere mentions a "cluster." The Office Action correctly acknowledges that *Rastogi et al.* does not explicitly teach cluster analysis "based on distances," and then relies on *Shimoji et al.* as disclosing "a method of clustering a set of data by using a clustering error based on distances between the data and respective one or more centers of the one or more clusters" (p. 6) *Shimoji et al.* is directed to clustering data based on entropical scheduling, where the assignment of a cluster to each data, for the update of the cluster center, is probabilistic, where the probabilities that each data belongs to individual clusters depend on the distances to the corresponding cluster centers (Abstract). Nowhere does

Shimoji et al. disclose or suggest "performing a cluster analysis <u>along the selected feature</u> to group the data into one or more clusters based on distances between the data and respective one or more centers of the one or more clusters." In fact, the data of *Shimoji et al.* is defined over a d-dimensional space, and the clustering error is "measured by the Euclidean distance" in d-space, (Introduction, page 2423, right column) and thus there is no suggestion for a cluster analysis "along the selected feature."

As motivation for a combination of *Rastogi et al.* in view of *Shimoji et al.,* the Examiner contends, "to combine clustering error as taught by Shimoji to analyze a cluster when grouping data into one or more cluster of a decision tree." However, the Examiner fails to explain how one skilled in the art would utilize the "clustering error" of *Shimoji et al.* (Equation (1), page 2423) in combination with *Rastogi et al.*, which nowhere even mentions "clusters," much less any "distances" between any data and other objects. In fact, even if *Rastogi et al.* had any clusters, any type of added "cluster analysis" would be technically infeasible, as *Rastogi et al.* already discloses an equation for entropy for a set of records, based on relative frequencies of respective classes in the set (e.g., "the more homogeneous a set is with respect to the classes of records in the set, the lower is the entropy"), and an equation for entropy of a split to divide the set, and states, "Consequently, the split with the least entropy best separates classes, and is thus chosen as the best split for a node." Thus, there is no motivation to combine *Rastogi et al.* and *Shimoji et al.*, other than impermissible hindsight. Thus, the rejection of claims 1-3 and 18-22 based on *Rastogi et al.* in view of *Shimoji et al.* should be withdrawn.

2. <u>CLAIMS 4 AND 21 ARE NOT RENDERED OBVIOUS BY *RASTOGI ET AL., SHIMOJI ET AL.*, AND *BACKGROUND*.</u>

The rejection of claims 4 and 21 based on *Rastogi et al.*, *Shimoji et al.*, and *Background* should also be reversed. The *Background* does not cure the factual deficiencies or the lack of motivation to combine *Rastogi et al.* and *Shimoji et al.* Therefore claims 4 and 21 are patentable over the applied art.

3. <u>CLAIMS 5 AND 22 ARE NOT RENDERED OBVIOUS BY *RASTOGI ET AL., SHIMOJI ET AL., BACKGROUND*, AND *HALL ET AL.*</u>

Hall et al. does not fix the problems with the rejection of claims 5 and 22 based on *Rastogi et al.*, *Shimoji et al.*, and *Background. Hall et al.* merely relates to a technique for generating pre-analyzed clusters for use in a conventional decision tree building algorithm. *Hall et al.* is directed to a method of developing of fuzzy rules from continuous valued data by building a decision tree in accordance with the C4.5 algorithm (Abstract, p. 1757, col. 1). However, *Hall et al.* recognizes that the "C4.5 algorithm tree algorithm **requires crisp** class assignments for all objects. It is **necessary** to partition the continuous output values into a effect set of **discrete** output classes." (Section 2.1, p. 1758, col. 1, emphasis added). *Hall et al.* thus proposes to preprocess the data initially by applying a fuzzy c-means clustering to determine the discrete classes, and then feeding the discrete classes into the C4.5 algorithm: "After a discrete class has been created for each example, as discussed in Section 2.1, C4.5 may be used to create a decision tree." (Section 3, p. 1759, col. 1). Therefore, whatever cluster analysis performed in *Hall et al.*, that cluster analysis must be performed **before**, not during, the building of the decision tree with the C4.5 algorithm. As a result, there is no teaching or suggesting in *Hall et al.*, of recursively performing the cluster analysis while "refining a node of a decision tree."

This lack of teaching is reason enough that there is not a factual basis to sustain the Examiner's rejection. In fact, what little disclosure of *Hall et al.*'s clustering method happens to **teach against** the recursively cluster analysis of the claims. *Hall et al.*'s C.4.5 algorithm requires crisp classes, and, if the classes are crisp at the beginning of the C4.5 algorithm, no cluster analysis during execution of the C4.5 algorithm would be necessary, teaching against "recursively … performing the cluster analysis on the projected data in each of the clusters" in a method for "refining a node of a decision tree."

4. CLAIMS 6 AND 23 ARE NOT RENDERED OBVIOUS BY *RASTOGI ET AL., SHIMOJI ET AL.*, AND *SHAFER ET AL.*

The rejection of claims 6 and 23 based on *Rastogi et al.*, *Shimoji et al.*, and *Shafer et al.* should also be reversed. As explained in Section VII.B.5, *Shafer et al.* does not the hard cluster analysis either.

VIII. CONCLUSION AND PRAYER FOR RELIEF

For the foregoing reasons, Appellants request the Honorable Board to reverse each of the Examiner's rejections.

Respectfully Submitted,

DITTHAVONG & CARLSON, P.C.

1/19/2005
Date

Margo Livesay, Ph.D.
Reg. No. 41,946

Stephen C. Carlson
Reg. No. 39,929

Attorneys for Appellant(s)

10507 Braddock Rd, Suite A
Fairfax, VA 22032
Tel. 703-425-8516
Fax. 703-425-8518

APPENDIX

1. (Previously Presented) A method for refining a node of a decision tree associated with a plurality of data characterized by a plurality of features, comprising:

selecting a feature from among the features characterizing the data associated with the node;

performing a cluster analysis along the selected feature to group the data into one or more clusters based on distances between the data and respective one or more centers of the one or more clusters;

constructing one or more arcs of the decision tree at the node respectively for each of the one or more clusters;

projecting the data in each of the clusters, wherein the projected data are characterized by the plurality of the features but for the selected feature; and

recursively performing the steps of selecting a feature and performing the cluster analysis on the projected data in each of the clusters.

2. (Original) The method according to claim 1, wherein the step of selecting the feature includes the steps of:

performing a plurality of cluster analyses along each of the features to calculate a maximal cluster validity measure, said maximal cluster validity measure corresponding to one of the features; and

selecting the one of the features that corresponds to the maximal cluster validity measure.

3. (Original) The method according to claim 2, wherein the step of performing a plurality of cluster analyses along each of the features to calculate a maximal cluster validity measure includes the performing the steps of:

for each of the features, performing a plurality of cluster analyses along said each of the features for a plurality of cluster numbers to calculate respective partition coefficients; and

determining the maximal cluster validity measure from among the partition coefficients.

4. (Original) The method according to claim 1, wherein the step of performing the cluster analysis includes the step of performing a fuzzy cluster analysis.

5. (Original) The method according to claim 4, wherein the step of performing the fuzzy cluster analysis includes the step of performing a fuzzy c-means analysis.

6. (Original) The method according to claim 1, wherein the step of performing the cluster analysis includes the step of performing a hard cluster analysis.

7. (Previously Presented) A method for refining a node of a decision tree associated with a plurality of data characterized by a plurality of features, comprising:

selecting a feature from among the features characterizing the data associated with the node;

performing a cluster analysis along the selected feature to group the data into one or more clusters;

constructing one or more arcs of the decision tree at the node respectively for each of the one or more clusters;

projecting the data in each of the clusters, wherein the projected data are characterized by the plurality of the features but for the selected feature; and

recursively performing the steps of selecting a feature and performing the cluster analysis on the projected data in each of the clusters,

wherein the step of performing the cluster analysis along the selected feature to group the data into one or more clusters includes the steps of:

calculating a domain ratio of a difference in domains limits of the data over a difference in domain limits of a superset of the data;

determining whether the domain ratio has a predetermined relationship with a predetermined threshold; and

if the domain ratio has the predetermined relationship with the predetermined threshold, then grouping the data into a single cluster.

8. (Original) The method according to claim 7, wherein the step of determining whether the domain ratio has the predetermined relationship with the predetermined threshold includes the step of determining whether the domain ratio is less than the predetermined threshold.

9. (Canceled)

10. (Previously Presented) A method for generating a decision tree for a plurality of data characterized by a plurality of features, comprising:

performing a plurality of cluster analyses along each of the features to calculate a plurality of respective partition coefficients based on membership functions of the data for one or more clusters in respective said cluster analyses;

selecting the one of the features corresponding to a maximal partition coefficient from among the partition coefficients;

subdividing the data into one or more groups based on the selected feature; and

building the decision tree based on the one or more groups.

11. (Canceled)

12. (Original) The method according to claim 10, wherein the step of performing the cluster analyses includes the step of performing a plurality of fuzzy cluster analyses.

13. (Original) The method according to claim 10, wherein the step of performing the fuzzy cluster analyses includes the step of performing a plurality of fuzzy c-means analyses.

14. (Original) The method according to claim 10, wherein the step of performing the cluster analyses includes the step of performing a plurality of hard cluster analyses.

15. (Original) The method according to claim 10, wherein the step of performing the cluster analyses includes the steps of:

calculating a domain ratio of a difference in domains limits of the data over a difference in domain limits of a superset of the data;

determining whether the domain ratio has a predetermined relationship with a predetermined threshold; and

if the domain ratio has the predetermined relationship with the predetermined threshold, then grouping the data into a single cluster.

16. (Original) The method according to claim 10, wherein building the decision tree based on the one or more groups includes the steps of:

projecting the data in each of the groups, wherein the projected data are characterized by the plurality of the features but for the selected feature; and

recursively performing the steps of selecting a feature, comprising selecting a new one of the features corresponding to a new maximal partition coefficient and subdividing the data into one or more new groups based on the selected new feature.

17. (Original) A method for generating a decision tree for a plurality of data characterized by a plurality of features, comprising:

performing a plurality of fuzzy cluster analyses along each of the features to calculate a maximal partition coefficient and a corresponding set of one or more fuzzy clusters, said maximal partition coefficient corresponding to one of the features;

selecting the one of the features corresponding to the maximal partition coefficient; and

building the decision tree based on the corresponding set of one or more fuzzy clusters.

18. (Previously Presented) A computer-readable medium bearing instructions for refining a node of a decision tree associated with a plurality of data characterized by a plurality of features, said instructions being arranged to cause one or more processors upon execution thereby to perform the steps of:

selecting a feature from among the features characterizing the data associated with the node;

performing a cluster analysis along the selected feature to group the data into one or more clusters based on distances between the data and respective one or more centers of the one or more clusters;

constructing one or more arcs of the decision tree at the node respectively for each of the one or more clusters;

projecting the data in each of the clusters, wherein the projected data are characterized by the plurality of the features but for the selected feature; and

recursively performing the steps of selecting a feature and performing the cluster analysis on the projected data in each of the clusters.

19. (Original) The computer-readable medium according to claim 18, wherein the step of selecting the feature includes the steps of:

performing a plurality of cluster analyses along each of the features to calculate a maximal cluster validity measure, said maximal cluster validity measure corresponding to one of the features; and

selecting the one of the features that corresponds to the maximal cluster validity measure.

20. (Original) The computer-readable medium according to claim 19, wherein the step of performing a plurality of cluster analyses along each of the features to calculate a maximal cluster validity measure includes the performing the steps of:

for each of the features, performing a plurality of cluster analyses along said each of the features for a plurality of cluster numbers to calculate respective partition coefficients; and

determining the maximal cluster validity measure from among the partition coefficients.

21. (Original) The computer-readable medium according to claim 18, wherein the step of performing the cluster analysis includes the step of performing a fuzzy cluster analysis.

22. (Original) The computer-readable medium according to claim 21, wherein the step of performing the fuzzy cluster analysis includes the step of performing a fuzzy c-means analysis.

23. (Original) The computer-readable medium according to claim 18, wherein the step of performing the cluster analysis includes the step of performing a hard cluster analysis.

24. (Previously Presented) A computer-readable bearing instructions for refining a node of a decision tree associated with a plurality of data characterized by a plurality of features, said

instructions being arranged to cause one or more processors upon execution thereby to perform the steps of:

selecting a feature from among the features characterizing the data associated with the node;

performing a cluster analysis along the selected feature to group the data into one or more clusters;

constructing one or more arcs of the decision tree at the node respectively for each of the one or more clusters;

projecting the data in each of the clusters, wherein the projected data are characterized by the plurality of the features but for the selected feature; and

recursively performing the steps of selecting a feature and performing the cluster analysis on the projected data in each of the clusters,

wherein the step of performing the cluster analysis along the selected feature to group the data into one or more clusters includes the steps of:

calculating a domain ratio of a difference in domains limits of the data over a difference in domain limits of a superset of the data;

determining whether the domain ratio has a predetermined relationship with a predetermined threshold; and

if the domain ratio has the predetermined relationship with the predetermined threshold, then grouping the data into a single cluster.

25. (Original) The computer-readable medium according to claim 24, wherein the step of determining whether the domain ratio has the predetermined relationship with the predetermined threshold includes the step of determining whether the domain ratio is less than the predetermined threshold.

26. (Canceled)

27. (Previously Presented) A computer-readable medium bearing instructions for generating a decision tree for a plurality of data characterized by a plurality of features, said instructions being arranged to cause one or more processors upon execution thereby to perform the steps of:

performing a plurality of cluster analyses along each of the features to calculate a plurality of respective partition coefficients based on membership functions of the data for one or more clusters in respective said cluster analyses;

selecting the one of the features corresponding to a maximal partition coefficient from among the partition coefficients;

subdividing the data into one or more groups based on the selected feature; and

building the decision tree based on the one or more groups.

28. (Canceled)

29. (Original) The computer-readable medium according to claim 27, wherein the step of performing the cluster analyses includes the step of performing a plurality of fuzzy cluster analyses.

30. (Original) The computer-readable medium according to claim 27, wherein the step of performing the fuzzy cluster analyses includes the step of performing a plurality of fuzzy c-means analyses.

31. (Original) The computer-readable medium according to claim 27, wherein the step of performing the cluster analyses includes the step of performing a plurality of hard cluster analyses.

32. (Original) The computer-readable medium according to claim 27, wherein the step of performing the cluster analyses includes the steps of:

calculating a domain ratio of a difference in domains limits of the data over a difference in domain limits of a superset of the data;

determining whether the domain ratio has a predetermined relationship with a predetermined threshold; and

if the domain ratio has the predetermined relationship with the predetermined threshold, then grouping the data into a single cluster.

33. (Original) The computer-readable medium according to claim 27, wherein building the decision tree based on the one or more groups includes the steps of:

projecting the data in each of the groups, wherein the projected data are characterized by the plurality of the features but for the selected feature; and

recursively performing the steps of selecting a feature, comprising selecting a new one of the features corresponding to a new maximal partition coefficient and subdividing the data into one or more new groups based on the selected new feature.

34. (Original) A computer-readable medium bearing instructions for generating a decision tree for a plurality of data characterized by a plurality of features, said instructions being arranged to cause one or more processors upon execution thereby to perform the steps of:

performing a plurality of fuzzy cluster analyses along each of the features to calculate a maximal partition coefficient and a corresponding set of one or more fuzzy clusters, said maximal partition coefficient corresponding to one of the features;

selecting the one of the features corresponding to the maximal partition coefficient; and

building the decision tree based on the corresponding set of one or more fuzzy clusters.

35. (Previously Presented) The method of claim 17 wherein the maximal partition coefficient is based on membership functions of the data for the set of one or more clusters.

36. (Previously Presented) The computer-readable medium of claim 34, wherein the maximal partition coefficient is based on membership functions of the data for the set of one or more clusters,